

Nonparametric Inference in Nonlinear Principal Components Analysis: Exploration and Beyond

Linting, Mariëlle,
Nonparametric Inference in Nonlinear Principal Components Analysis:
Exploration and Beyond
Dissertation Leiden University – With ref. – With Summary in Dutch.

Subject headings: Nonlinear principal components analysis; PCA;
CATPCA; optimal scaling; nonparametric inference; nonparametric
bootstrap; stability; permutation tests; statistical significance

ISBN 978-90-9022232-5

© 2007 Mariëlle Linting

Printed by Mostert & Van Onderen!, Leiden

Nonparametric Inference in
Nonlinear Principal Components Analysis:
Exploration and Beyond

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op dinsdag 16 oktober 2007

klokke 16.15 uur

door

Mariëlle Linting

geboren te Alphen aan den Rijn

in 1977

PROMOTIECOMMISSIE

Promotor Prof. dr. J.J. Meulman

Referent Prof. dr. H.A.L. Kiers, University of Groningen

Overige Leden
Prof. dr. M.H. van IJzendoorn
Prof. dr. W.J. Heiser
Prof. dr. P.J.F. Groenen
Prof. dr. L.W.C. Tavecchio

Contents

Acknowledgements	ix
1 General Introduction	1
1.1 Categorical Data Analysis in the Social and Behavioral Sciences	2
1.1.1 Optimal quantification	4
1.1.2 Nonlinear PCA as an exploratory technique	4
1.2 Inference in Principal Components Analysis	5
1.2.1 Methods for nonparametric inference	6
1.2.2 Linear PCA	7
1.2.3 Nonlinear PCA	8
1.3 Outline	9
2 Introduction to Nonlinear PCA	11
2.1 Introduction	12
2.2 The Method of Nonlinear Principal Components Analysis . . .	13
2.2.1 Category quantification	14
2.2.2 Nonlinear and linear PCA: Similarities and differences .	27
2.3 Nonlinear PCA in Action	30
2.3.1 Software	30
2.3.2 The ORCE data	31
2.3.3 Choice of analysis method and options	35
2.3.4 The nonlinear PCA solution for the ORCE data	37
2.3.5 Comparison of the nonlinear and linear PCA solution .	44
2.4 Discussion	46
3 Stability of Nonlinear PCA	49
3.1 Introduction	50
3.2 Assessing Stability of Nonlinear PCA	52
3.2.1 The nonparametric bootstrap procedure	53
3.2.2 Validity of the bootstrap in nonlinear multivariate analysis	53

3.2.3	The bootstrap procedure in nonlinear PCA	55
3.2.4	Confidence ellipses	59
3.3	Application	60
3.3.1	The Observational Record of the Caregiving Environment (ORCE)	61
3.3.2	Balanced bootstrap results for the nonlinear PCA solution	64
3.3.3	A solution to the instability problem: Merging categories with small marginal frequencies	68
3.3.4	Comparing nonlinear PCA to linear PCA	76
3.4	Conclusions and Discussion	79
4	Permutation tests in linear PCA	85
4.1	Introduction	86
4.2	The Use of Permutation Tests in PCA	88
4.2.1	Two permutation strategies	89
4.3	Design of the Monte Carlo Study	90
4.3.1	Generating data matrices with different principal component structures	91
4.3.2	Correction for multiple testing	94
4.3.3	Computing proportions of Type I and Type II error	95
4.3.4	Choosing the number of Monte Carlo samples	96
4.4	Results	97
4.4.1	Permutation strategies: Overall comparison	97
4.4.2	Permutation strategies combined with different confidence level conditions	99
4.5	Conclusions and Discussion	105
5	Permutation tests in nonlinear PCA	111
5.1	Introduction	112
5.2	Permutation Tests	114
5.2.1	Permutation tests in linear PCA	115
5.2.2	Permutation tests in nonlinear PCA	116
5.3	Effect Size	116
5.4	Relation between Statistical Significance and Stability	118
5.5	Application: The ORCE Data	119
5.5.1	P -values for the contribution of the ORCE variables	122
5.5.2	Effect sizes for the ORCE variables	125
5.5.3	Significance of the contribution of the ORCE variables to the nonlinear PCA solution	128
5.6	Permutation and Bootstrap Results Compared	129
5.7	Conclusions and Discussion	133

6 General Discussion	137
6.1 A Short Retrospect	138
6.1.1 Nonlinear PCA as an exploratory method	138
6.1.2 Stability of nonlinear PCA	139
6.1.3 Statistical significance of the contribution of variables to the linear PCA solution	141
6.1.4 Statistical significance of the contribution of variables to the nonlinear PCA solution	142
6.2 Ideas for Further Research	144
6.2.1 The bootstrap	144
6.2.2 Permutation tests	145
6.3 Implementation	147
Appendix A The Mathematics of Nonlinear PCA	149
Appendix B Missing Data	155
Appendix C Construction of Confidence Ellipses	157
Appendix D Simulating Data With a Specific Component Structure	159
Appendix E Confidence Intervals for Type I and Type II Errors	161
References	163
Summary in Dutch (Samenvatting)	173
Samenvatting	174
NLPCA als exploratieve methode	175
Stabiliteit van NLPCA	176
Statistische significantie van de bijdrage van de variabelen aan de lineaire PCA-oplossing	178
Statistische significantie van de bijdrage van de variabelen aan de NLPCA-oplossing	180
Ideeën voor toekomstig onderzoek	182
De bootstrap	182
Permutatietoetsen	183
Implementatie	184
Curriculum Vitae	185

Acknowledgements

First, I would like to thank the people whose useful suggestions and comments have greatly contributed to this thesis: Louise Fitzgerald for combining her technical and verbal skills to improve the wording of Chapter 2; Lawrence Hubert, Paul Eilers, and Georgi Nalbantov for carefully reading and annotating Chapter 3; Bart Jan van Os for the many discussions about the content of Chapter 4 and his continuing advice on programming in Matlab; and a still extending number of anonymous reviewers for their elaborate and constructive remarks.

The data set that is repeatedly used in this thesis to illustrate the procedures was collected by the NICHD Early Child Care Research Network, supported by the NICHD through a cooperative agreement. I acknowledge the generous way in which the NICHD Study on Early Child Care has made this unique data set available for secondary analyses.

And last, but not least, there are some people I would like to thank for simply being there. And as I will probably never get an Oscar, I will take this opportunity to do so: I thank my dear colleagues and friends Anita (for her help with completing this thesis and the cynicism and fun during our delicious Thursday night dinners), Elise (for the emotional support and good advice), Willy (for the therapy sessions), and Ellen (for the small talk and subjecting the final version of this thesis to her artistical skills). Also, great thanks to my friends Emily, Caroline, Jantina, Brechje and Gabriëlle for reminding me of life beyond work; my parents who endured some occasional fury, even when they had no idea what it was about; John and Nora for their continuous support; and, of course, to Sander whose “Katsepia” helped me put everything into perspective.

